Paul Tempst
Andrew J. Link
Lise R. Riviere
Mark Fleming
Christopher Elicone

Howard Hughes Medical Institute,
Department of Genetics,
Harvard Medical School,
Boston, MA

Internal sequence analysis of proteins separated on polyacrylamide gels at the submicrogram level: Improved methods, applications and gene cloning strategies

The fields of protein chemistry and molecular biology are currently merging for study of biologically relevant events and conditions. To obtain partial sequences of microamounts of protein, efficient integration of high resolution separation and sequencing technologies is required. We report here on improved methods that allow extensive internal sequencing of 10 to 20 picomoles protein recovered from one- or two-dimensional gels. Each step of the standard protocol of Aebersold *et al.* (*Proc. Natl. Acad. Sci. USA* 1987, *84*, 6970–6974) and the required instrumentation were examined and specifically adapted for use with submicrogram amounts of protein. Optimizations of *in situ* microdigests and liquid chromatography were needed for improved peptide recovery. Subsequent automated sequencing required subpicomole analysis. New methods for S-alkylation of gel-separated proteins and accurate identification of tryptophan-containing peptides were introduced to insure overall higher efficiencies. The acquired internal sequences facilitated cloning of the genes and several strategies are discussed. Applying our method, several proteins of unknown structure were sequenced and successfully identified or cloned. Internal sequences of submicrogram protein amounts, recovered from a single two-dimensional gel of *Escherichia coli* total protein (120 μg), allowed unambiguous identification of the spots but pre-gel enrichment will be required for analysis of most (90–95 %) other spots. Integration of comprehensive two-dimensional gel protein databases with methods and strategies outlined here could potentially be an abundant source of DNA probes and markers useful for guidance of the human genome sequencing project and for analysis of the emerging vast amounts of data.

1 Introduction

In recent years, many biological research projects were a cooperative approach between protein chemistry and molecular biology. The major practical integration of the two fields is the use of partial protein sequence for the design of oligonucleotide probes to clone the corresponding gene [1]. The cloned genes of low-abundant proteins could be easily sequenced and used for *in vitro* mutagenesis or as a probe in the study of transcriptional regulation. Analysis of the proteins is crucial to obtain information on posttranslational modifications and processing, domain structures, surface topographies and folding. One limiting factor of this combined method is the sensitivity of protein sequencing technology. Although some disagreement exists on the exact amount, approximately 5 picomoles of purified protein is currently required for analysis [2–4]. By default, any protein of great importance will only be available at quantities just below the required minimum level after complete purification. Quite often, most of the precious material was lost during final purification, concentration or attempted removal of buffer components that interfere with chemical protein sequencing (*e.g.* salts, detergents, amines). During the last few years, direct sequencing of proteins separated on polyacrylamide gels and electroblotted onto a suitable solid support has been introduced [5–8]. Since this method allowed efficient integration of a high-resolution, micropreparative separation technique, volume reduction and removal of interfering substances with automated sequencing, it has found widespread use and has been intensively studied and optimized [4, 9, 10]. Picomolar quantities of partially purified proteins could be resolved on one-dimensional electrophoresis (1-DE) gels and the protein of interest sequenced.

The combination of two-dimensional gel electrophoresis (2-DE) with direct sequencing has also been reported [10–12] and may turn out to be an important research tool in the future. Since its introduction [13–14], high resolution 2-DE has been frequently used to analyze constitutive or induced variations between cells or tissues. The technique allowed examination of cellular phenotypes, at the single protein level, in studies on differentiation, second messengers, transmitters, growth factors, neoplastic transformation, toxicological effects or genetic diseases [15–17]. Computer analysis of the 2-DE patterns [18] was used to construct comprehensive databases for quantitative and descriptive data on cellular proteins that showed changing levels of expression [19]. Direct sequencing could provide a straightforward and positive identification of these proteins or yield sequence information to clone the genes.

Although extensive precautionary measures have been implemented [10], amino terminal blocking of ultramicro amounts of proteins during 2-DE has generally been observed, rendering them unsuitable for direct sequence analysis. The

Correspondence: Dr. Paul Tempst, Department of Genetics, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

Abbreviations: **1-DE** and **2-DE**, one- and two-dimensional electrophoresis; **AUFS**, absorption units full scale; **BSA**, bovine serum albumin; **CAPSO**, (3-[cyclohexylamino]-2-hydroxyl-1-propanesulfonic acid); **DTT**, dithiothreitol; **GUHCl**, guanidine hydrochloride; **IEF**, isoelectric focusing; **kb**, kilobase pairs; **kDa**, kilodalton; **MeCN**, acetonitrile; **NC**, nitrocellulose; **PAGE**, polyacrylamide gel electrophoresis; **PCR**, polymerase chain reaction; **PE-Cys**, S-β-(4-pyridylethyl) cystein; **PTH**, phenylthiohydantoin; **PVDF**, polyvinylidene difluoride; **PVP-40**, polyvinyl pyrrolidone, average M_r 40 000; **RP-HPLC**, reverse phase – high performance liquid chromatography; **SDS**, sodium dodecyl sulfate; **TFA**, trifluoroacetic acid; **TIM**, triose phosphate isomerase

0173-0835/90/0707-0537 $3.50+.25/0

EXHIBIT I

same applies for the many cellular proteins that have naturally blocked *N*-termini due to posttranslational modification. In the absence of an adequate remedy, internal sequence analysis is the only alternative. This requires cleavage of the protein and separation of the resulting peptides. Different techniques, specifically for proteins separated on gels, have been developed for this purpose [20–22]. Multiple stretches of acquired internal sequence allow for more accurate identification of the protein [20] and facilitates cloning of the gene by either conventional or polymerase chain reaction (PCR) technologies [23, 24].

Poor yields are a recurring problem associated with micropreparative digests and separations. Therefore, either more sensitive sequence analysis techniques or more starting material will be required. If this rechnique is ever to become a more general research tool in the biological sciences, one must be able to cope not only with the most abundant proteins but also with the smallest spots. Unfortunately, 2-DE gels have a limiting loading capacity and minimal protein mass is preferable for optimal resolution. Different estimates have appeared in the literature of the minimal amount of a single protein on a 2-DE gel required for internal sequencing, and the maximum number of cellular proteins that can be sequenced without a pre-gel enrichment. Most of these claims are largely exaggerated; meticulous inspection of several published experiments revealed that end-point analysis often indicated even larger quantities than the presumed starting amounts.

We have recently been involved in several collaborative studies that required internal sequence analysis of proteins separated on 1-DE gels. In addition, we conducted a detailed study on the practical aspects and limitations of dealing with proteins separated on 2-DE gels. The feasibility of internal analysis of low-picomolar quantities, with maximal sequence information (no gaps) was investigated in particular. This report describes the results and the implementation of corrective measures for improved efficiency of the different steps leading to, and including, micro sequence analysis and gene cloning.

2 Materials and methods

2.1 Materials

All standard proteins were purchased from Sigma (St. Louis, MO). Yeast 6-phosphofructo-2-kinase (6PF2K) was obtained from Dr. M. Kretschmer, calcium activated serine protease from brain (CASP) from Dr. C. Abraham, phage SP6 DNA polymerase from Dr. J. Rush, inositol triphosphate receptor (I3PR) from Dr. C. Chadwick and brain protein BD-43 from Dr. B. Denker; all proteins, except the I3PR, were only partially purified when received. Test peptides were synthesized in our laboratory using solid-phase procedures. Acrylamide, *N,N'*-methylenebisacrylamide, sodium dodecyl sulfate (SDS), urea and Resolyte pH 4–8 carrier ampholytes were obtained from BDH (Poole, UK); Servalyt, pH 8–10 carrier ampholytes were from Serva (Heidelberg, FRG), dithiothreitol (DTT) was from Calbiochem (La Jolla, CA), Nonidet P-40 (NP-40) and beta mercaptoethanol from Sigma and *m*-Cresol Purple from US Biochemical Corporation (Cleveland, OH). All other materials are listed elsewhere in this section.

2.2 Protein extracts of *Escherichia coli* EM 62

Escherichia coli strain EMG2 (ATCC 23716) was grown to stationary phase in complete media (1 % tryptone, 1 % yeast extract, 0.5 % NaCl, pH 7.5) at 37 °C. A protion (400 µL) of the culture was centrifuged and washed with 10 mM $MgCl_2$, 1 mM Tris-HCl, pH 7.4. The cells were gently resuspended in 10 µL of lysis buffer (50 mM Tris-HCl, pH 6.8, 2 % SDS, 5 % v/v glycerol, 5 % v/v β-mercaptoethanol) and heated at 95 °C for 4 min. The lysate was quickly cooled to room temperature and 90 µL of solubilizer solution (9 M urea, 4 % NP-40, 2 % Servalyt pH 8–10 carrier ampholytes, 1 % DTT) were added. The protein extract was immediately loaded onto isoelectric focusing (IEF) gel or stored frozen at –70 °C. Volumes varying from 2.5 µL (5–7 µg) to 30 µL (50–75 µg) of protein extract were used for 2-DE analysis. The protein concentration was assayed using a modified Bradford assay [25].

2.3 Radioiodination

Iodinations of protein mixtures were done using Iodo-beads (Pierce, Rockford, IL) and Na ^{125}I (Amersham, Arlington Heights, IL) according to the manufacturer's instructions and as described [26]. Proteins were separated from unincorporated label either on a Sephadex G-25 column (Pharmacia, Piscataway, NJ) or on a 1-DE gel. Only a few percent of the mixtures were labeled and used to spike the bulk of the protein mass or for separate experiments.

2.4 Polyacrylamide gel electrophoresis

One-dimensional SDS-polyacrylamide gel electrophoresis (PAGE) was done essentially as described by Laemmli [27]. Gels (4 % stacking gel, varying % running gels) were usually 9 cm × 8 cm × 0.5 mm (thick) in size, with 6 × 9 mm wells, and were run at 15 mA constant current (30 mA for 1 mm thick gels) at room temperature. Proteins, maximally 50 µg per well, were heated in sample buffer for 10 min at 60 °C before loading on the gel. 2-DE was performed using a modification of the method described by Anderson [28]. First-dimensional IEF (1h at 200 V, prefocusing; 14h at 800 V, separation) was in 150 × 1.5 mm, 3.3 % polyacrylamide gels, containing 2 % w/v carrier ampholytes (Resolyte pH 4–8), simultaneously carried out for 10 tubes, in an Iso-Dalt tank (LSB Corporation, Rockville, MD) at room temperature. The IEF rod gel was then soaked for 20 min at ambient temperature in 10 mL equilibration solution (10 % v/v glycerol, 2 % w/v SDS, 8.6 mM DTT, 0.125 M Tris-HCl, pH 6.8, 0.1 % w/v *m*-Cresol Purple), immediately transferred to the second dimension and secured in position with 1 mL of agarose solution (0.5 % w/v agarose dissolved in 10 % v/v glycerol, 2 % w/v SDS, 0.125 M Tris-HCl, pH 6.8, 8.6 mM DTT). Alternatively the first-dimensional gel was stored in equilibrium solution at –70 °C until needed. Second-dimensional SDS-PAGE (160 × 155 × 1.5 mm; 12.5 % T), was simultaneously carried out for 4 slabs, in a model SE600 Hoeffer (San Francisco, CA) tank, at 20 °C and 120 mA constant current (30 mA per gel) until the tracking dye had run off the lower edge of the gel (usually 7h). The Laemmli anode buffer was supplemented with 50 mM sodium acetate.

2.5 Coomassie Brilliant Blue staining

Staining with Coomassie Brilliant Blue R-250 was as described by Anderson [29]. Briefly, gels were fixed in 50 % ethanol,

2 % H_3PO_4 for 2 h, washed three times for 20 min with water and equilibrated for 3 h with 34 % methanol, 17 % ammonium sulfate, 3 % H_3PO_4. Powdered Coomassie Brilliant Blue R-250 (Bio-Rad, Richmond, CA) was then added (0.5 g/100 mL) and staining carried out for 3 days on a shaker platform. The gels were washed with water for 30 min and dried for storage using the method of Samal [30].

2.6 Electroblotting and staining

Transfer of proteins from 1-DE gels to nitrocellulose (NC) membranes (Schleicher & Schull, Keene, NH) [31] was done in 25 mM Tris, 192 mM glycine, 20 % v/v methanol buffer for 15 h at 30 V constant voltage in a transblot tank from Bio-Rad at 10 °C. For 2-DE gels, transfers were done in 25 mM (3-[cyclohexylamino]-2-hydroxy-1-propanesulfonic acid) CAPSO (Sigma) buffer, 20 % v/v methanol for 15 h at 30 V [32]. Blots from all 1-DE gels and micropreparative (> 30 µg total protein per gel) 2-DE gels were stained for 1 min with 0.1 % Ponceau S (Fluka, Buchs, Switserland) solution in 1 % acidic acid and then washed with 1 % acetic acid for 1 min [20]. Bands or spots were then cut with a razor blade while submerged in water (in a Petri dish) and stored wet at –20 °C in an Eppendorf tube, containing 500 µL MilliQ water. Blots from analytical 2-DE gels were gold stained with Aurodye Forte (Janssen Life Sciences, Beerse, Belgium) according to the manufacturer's instructions. Each blot (16 × 20 cm) was incubated, during the last step, with 50 mL dye in a heat-sealed plastic bag for 3 h, washed with water and air-dried.

2.7 *In situ* digests

Digestion of NC-bound proteins was done using a modification of the method described [20], as follows. Frozen spots or bands were thawed and destained with 1 mL 0.2 mM NaOH for 1 min followed by 2 rinses with water and then incubated for 30 min at 37 °C with 1 mL 0.5 % polyvinyl pyrrolidone (PVP-40, Sigma) in 100 mM acetic acid followed by 5 rinses with water, all done in a 1.5 mL Eppendorf tube. NC strips were then cut into small squares (1 × 1 mm) while submerged in water. The excess liquid was drained by squeezing all the NC-pieces together with a fine-tipped forceps. Pieces were then transferred to a 500 µL Eppendorf tube containing a minimal volume of 100 mM NH_4HCO_3/acetonitrile (95:5 v/v). The protease of choice was added in a concentration of > 0.04 µg/µL; enzyme/substrate ratios varied from 1/10 (for 10 µg substrate) to 2 (for 0.5 µg substrate). Incubation was always done for 15 h at 37 °C. Mild vortexing was done after every change in incubation or washing solution, during the entire procedure, to ensure exposure of all surfaces; small pieces of NC tend to stick together. After the digest, the suspension was sonicated for 5 min, NC was spun down for 1 min at 15 000 rpm in a Eppendorf centrifuge and the supernatant transferred to another tube. The NC pieces were rinsed once with an equal volume of digest buffer. Combined supernatants were immediately injected for reverse phase-high performence liquid chromatography (RP-HPLC) analysis or stored frozen at –20 °C. Trypsin, *Staphylococcus aureus* V8 endoproteinase Glu-C and *Pseudomonas fragii* endoproteinase Asp-N were all "sequencing grade" from Boehringer (Indianapolis, IN), *Achromobacter lyticus* endoproteinase Lys-C (lysyl endopeptidase) was from Wako Chemicals (Osaka, Japan) and chymotrypsin and subtilisin were from Sigma.

2.8 Reduction and *S*-alkylation

Reduction and S-pyridyl ethylation were done either before the gels were run or after the *in situ* digest, prior to RP-HPLC.

2.8.1 Pre-gel alkylation

Reduction was carried out in Laemmli sample buffer [27], containing 0.5 % β-mercaptoethanol as reducing agent, for 10 min at 60 °C followed by 20 min at 37 °C. A 20 % v/v solution of 4-vinylpyridine (Sigma) in ethanol was then added to yield a final concentration of 1.5 %. The alkylation reaction was allowed to proceed for 30 min at room temperature in the dark followed by immediate loading on the gel and starting of the electrophoretic run.

2.8.2 Post-digest alkylation

β-Mercaptoethanol was added to the combined supernatants, after *in situ* digests (see Section 2.5), at a final concentration of 0.1 % and reduction carried out for 30 min at 37 °C. 4-Vinylpyridine (20 % solution in ethanol) is added at a final concentration of 0.3 % and the reaction is carried out for 30 min at room temperature in the dark. The mixture was then immediately injected for HPLC separation.

2.9 RP-HPLC

The system used for narrow-bore liquid chromatography in this study consisted of a Microgradient System (Brownlee Labs, Santa Clara, CA) equipped with a 200 µL dynamic mixer; a 3.2 × 15 mm RP-4 Newguard column (ABI, Foster City, CA) was inserted between the mixing T and the dynamic mixer. Samples were loaded using a Rheodyne model 7125 injector, obtained from Rainin (Woburn, MA) with a 100 µL loop. All precolumn plumbing was done with 0.007 inch ID, stainless steel tubing (Upchurch, Oak Harbor, WA) with a combined dead volume of 30 µL. Total precolumn dead volume (from the mixing T), including sample loop, was 460 µL. The column outlet was directly connected with an uninterrupted piece of 0.005 inch ID tubing (2.5 µL dead volume) to the flow cell of either a 783 variable wavelength detector or a 1000 S diode-array detector (both from ABI, Ramsey, NJ). The latter was outfitted with a prototype 3 µL microflow cell (on loan from ABI for evaluation). The flow cell outlet line was identical to the inlet tubing (18 cm × 0.005 inch). Analog signals from the detectors were registered using one or two models SE120 (BBC Metrawatt/Goerz, Vienna, Austria) 2-channel stripchart recorders (4 channels total with the 1000S diode-array detector); in parallel, digitally converted signals were acquired on a PE Nelson (Cupertino, CA) datasystem using the Turbochrom (version 2700) software. A 2.1 × 220 mm Aquapore RP-300 column (ABI) was used during this study, except where indicated. Columns were operated at ambient temperature with a flow rate of 100 µL/min; the gradient was typically 15 min isocratic at 5 % B, linear 5–50 % B in 45 min, 50–100 % B in 20 min, except where indicated. Solvent A was 0.1 % trifluoroacetic acid (TFA, Pierce); solvent B was 0.09 % TFA in 70 % acetonitrile (Burdick & Jackson, Muskegon, WI). Manual fraction collection was as described in the text (Section 3.4.1). Fractions were never concentrated or dried and were stored at –20 °C until sequencing.

2.10 Peptide sequencing

Purified peptides were sequenced with the aid of an Applied Biosystems model 477A automated sequenator, operated according to the principles outlined by Hewick *et al.* [33]. Stepwise liberated phenylthiohydantoin amino acids were identified using an on-line 120A HPLC system, equipped with a PTH C18 (2.1×220 mm; 5 micron particle size) column (ABI). The standard ABI method was optimized for subpicomole PTH analysis as described [3].

2.11 Database searches

Searches for identical or homologous amino acid sequences in the Protein Identification Resource of the National Biomedical Research Foundation (Washington, DC) were done using the software package of the Genetics Computer Group (University of Wisconsin, Madison, WI).

3 Results and discussion

Different strategies to obtain internal sequence from proteins separated on 1-DE or 2-DE have been reported in the literature. Proteins were either electroblotted onto NC and digested *in situ*, and the resulting peptides were separated by HPLC followed by automated sequencing [20]; or proteins can be digested *in situ* in the polyacrylamide gel matrix following staining and destaining and the resulting peptides are either extracted, separated by HPLC and sequenced [21] or separated on another 1-DE gel followed by electroblotting onto polyvinylidene difluoride (PVDF) and direct sequencing of the bands [10, 22]. In our hands, the first method has worked at the highest levels of sensitivity and we have adopted it for routine work. Here we will describe our practical experiences, detailing versatility, efficiency and reproducibility of the technique and successful efforts to scale down to protein amounts in the 10–20 picomole range. Optimizations of microdigests, high sensitivity liquid chromatography and automated sequencing were required and will be discussed. Finally, simple steps were integrated into the existing protocol for *S*-alkylation of cysteines.

3.1 Electrophoresis and electroblotting

Since their introduction, 1-DE and 2-DE [13, 27] and electroblotting [31] techniques have been extensively used and optimized. All work in our laboratory has been done using these standard protocols. As the integration with internal sequencing is fairly new [20], a number of questions and problems need to be addressed: (i) are any special precautions required, (ii) what criteria can be used to decide whether enough protein is present, (iii) how many spots from a 2-DE gel can be readily sequenced, and (iv) how does one scale up? In this section, we will discuss results that may provide some of the answers. It should be understood that the technical improvements outlined in Sections 3.2–3.5 are a prerequisite for validity and reproducibility of our findings.

3.1.1 1-DE Gels

A number of partially purified proteins were separated on 1-DE gels and analyzed for internal sequence (results shown in Table 3). Only 2 μg (20 picomoles) or less of proteins PF2K-96K, PF2K-93K and SP6DP were available, as estimated by comparison of the Ponceau S-stained bands with several standards. Given the losses that were likely to occur at each step in the entire procedure, the initial sequencing yields (1.5–3 picomoles) of the derived peptides fit these estimates. All peptides were sequenced successfully and it is therefore safe to say that 0.2–2 μg of protein (10–100 kDa) on an NC blot should generally be sufficient for internal sequence analysis. Since no more than 50 μg total protein should be loaded in one well (1 cm) of a 1 mm thick minigel, the protein of interest must represent at least 0.5–2 % (M_r depending) of the total protein mass. Unlike amino-terminal sequencing, no precautions were needed during electrophoresis and electroblotting. Side chain modifications or destruction of Lys, Trp, Met or any other amino acid did not occur, as evidenced by the sequencing results (discussed in Section 3.5).

3.1.2 2-DE Gels

High resolution 2-DE and automated sequencing are specialized techniques that most often reside in different locations. A combination of the technologies usually requires shipment of samples; many problems and failures can be traced back to exactly this step. We therefore set out to implement the entire procedure in our laboratory. We opted for a modified Andersen 2-DE gel technique [28] that allows 2–20 gels to be run simultaneously. No special precautions against chemical modification were taken except that the second dimension gels were usually one day old. Using 1.5 mm thick 1-DE gels and a dilution series of 6 standard proteins, the sensitivity of different staining methods was tested. Either gels or NC-electroblots were stained. The following mimimal quantities could be visualized on a 2 x 2 mm square (average size of a 2-DE gel spot): 100 ng for Coomassie Brilliant Blue, 100 ng for Ponceau S and 5 ng for gold. A 2-DE pattern of 10 μg total protein of *E. coli*, electroblotted and stained with gold, reveals approximately 500 spots (Fig. 1A). In comparison, a 30 μg load gives 65 spots after Coomassie staining of the gel (Fig. 1B). Ponceau S staining of a blot revealed 57 and 80 spots for 30 and 75 μg loads, respectively (results not shown).

As discussed in Section 3.2.2, only limited amounts of NC are preferable for *in situ* microdigests of proteins; therefore not more than 6–12 spots (25 mm^2 combined) should be pooled in 25 μL of solvent. Since the minimal quantity of Ponceau-stained protein visible on a 4 mm^2 area is about 50–100 ng, pooling will maximally yield only 0.6 μg for the weakest spots. Using our method and the technologies described in the current report, this would be sufficient for analysis of most visualized proteins with molecular masses < 40 to 50 kDa. Three such protein spots of major intensity (Fig. 1, spots 1–3) were cut out from a blot (120 μg total protein) and internal sequence was obtained (Table 3, *E. coli* 2D 1, 2, 3). We estimate that by loading 200 μg per gel and pooling 10 spots, the number of sequenceable proteins will be approximately 30 to 40. Increasing this number further would depend solely on scaling up the total protein load. Whether trading increased loads for a general loss of resolution is justifiable will have to be determined on a case-by-case basis. We have no data on eukaryotic cellular proteins at present.

An elegant alternative to limit the amount of solid support (NC, PVDF or other) but still use protein spots from a theoretically unlimited number of gels has been described by Celis, Vandekerckhove and co-workers [34, 35]. Spots from

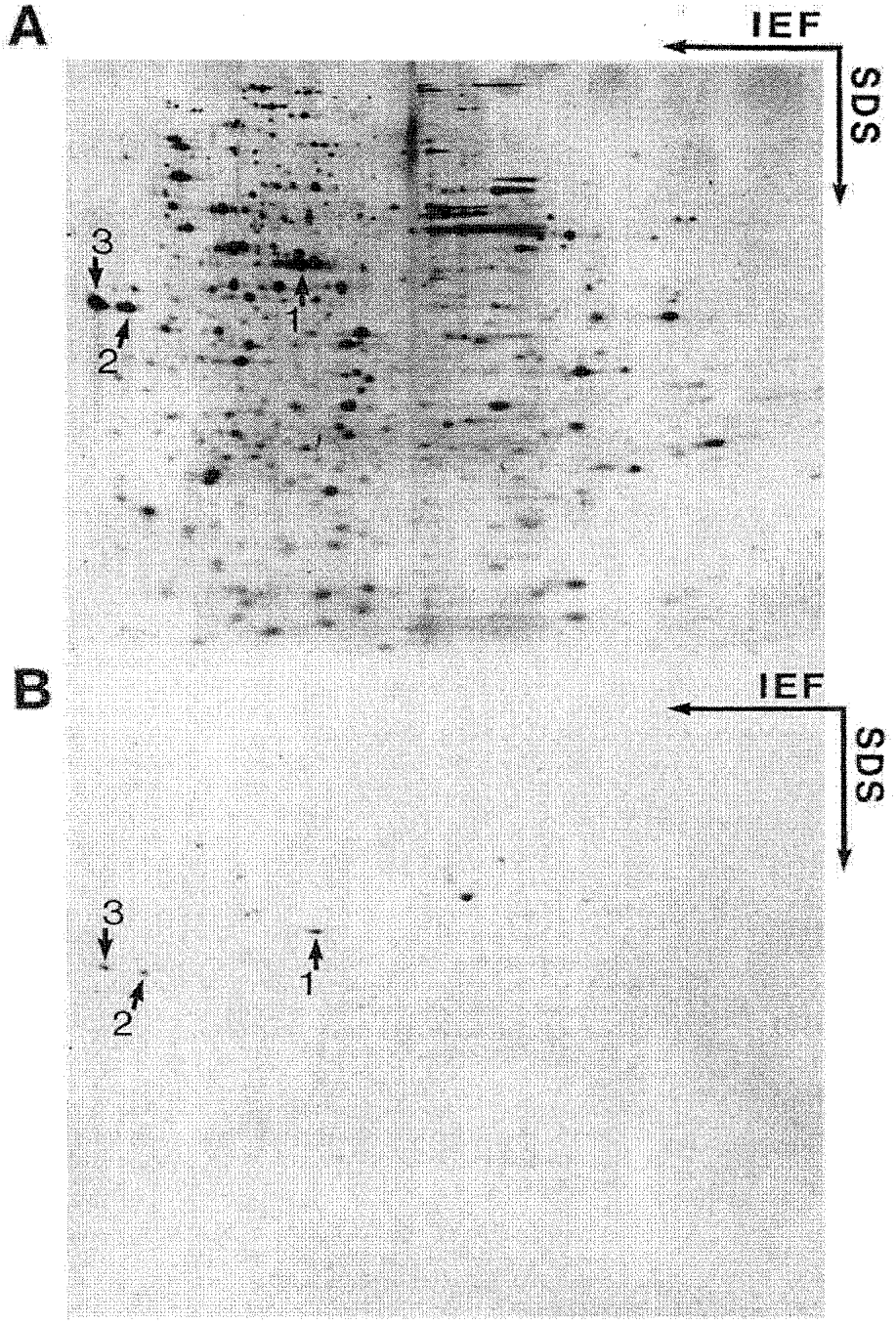

Figure 1. Total protein extract of *E. coli* separated on a 2-DE gel. A pH 4–8 carrier ampholyte gradient was used for the first dimension, gels were 12.5 % acrylamide in the second dimension. (A) 10 µg loaded; electroblotted onto NC, stained with Aurodye forte. (B) 30 µg loaded, gel stained with Coomassie Brilliant Blue. Arrow indicates spots that were analyzed for internal sequence.

Coomassie-stained 2-DE gels are placed in the well of a 1-DE SDS gel, electroeluted *in situ* and focused into one band in the stacking gel, followed by an electrophoretic run and an electroblot onto the support of choice. The sensitivity of the Coomassie staining is the only limiting factor of this method; details about yields were not available. However, this method provides the advantage that dried gel spots can be mailed from a gel laboratory to a sequencing facility where they are reswollen and processed. For the above two methods the detection of protein spots by either Ponceau or Coomassie staining is a prerequisite. If the protein of interest is not visible with either stain, it must be enriched in the total protein mass before loading on the focusing gel. This has been done by preliminary cell fractionation before making the protein extract (*e. g.* cytosol, cell membrane or mitochondrial proteins) and/or by different sorts of affinity chromatography [10, 36]. Alternatively, one may resort to classical liquid chromatography (*e. g.* ion exchange or size exclusion).

3.2 *In situ* digest

Our internal sequencing method of choice consists of three consecutive steps: *in situ* proteolysis, micropreparative liquid chromatography and automated amino acid sequencing. The original paper by Aebersold *et al.* [20], describing this technique, contains a highly optimized protocol for the protein transfer, stain and *in situ* digest. Systematic investigation in our laboratory of several experimental parameters, with regard to recovery and general applicability, indicated that almost none could be improved. In this section we describe our experiences with a variety of substrates, different proteases and the digest process in general, including optimization and practical guidelines for routine experiments at the low picomole level.

3.2.1 General technique

3.2.1.1 Substrates

As has been observed by others as well as us, a multitude of proteins, when undenatured and in solution, are resistant to digestion with most proteases, except for pepsin and proteinase K. Two notorious, proteolysis-resistant substrates are ribonuclease and triose phosphate isomerase (TIM). Until now, we have not come across any protein that could not be digested *in situ* to some extent. NC-bound TIM is successfully cleaved by trypsin, but also with endoprotease Lys-C, chymotrypsin, subtilisin and endoprotease Asp-N. Comparative peptide maps of TIM, using the first four proteases, are shown in Fig. 2. When comparing peak heights with digests of equimolar amounts of bovine serum albumin (BSA) and other standard proteins, we noticed that all peaks of the TIM digests were more than twofold lower, indicating that the digests may be incomplete. Moreover, endoproteinase Glu-C fails to digest NC-bound BSA whereas *S*-alkylated BSA, also NC-bound, is successfully cleaved by the same enzyme (result not shown). There are additional reasons to believe that the protein substrates are sometimes not completely reduced and denatured after Laemmli type electrophoresis and electroblotting (discussed in Section 3.3). Complete digests of TIM in solution can be easily obtained following guanidinium-hydrochloride-promoted denaturation (Riviere and Tempst, unpublished observations). However, guanidine hydrochloride (GuHCl) is incompatible with SDS gels and *in situ* GuHCl treatment of NC-bound proteins resulted in unacceptable washout. Improved results were obtained by reducing and alkylating proteins as described in Section 2 and in Section 3.3 (cysteine derivatization).

3.2.1.2 Proteases

For many applications described in the literature, trypsin was the enzyme of choice for *in situ* digestion of proteins. However, chymotrypsin, subtilisin, endoproteinases Lys-C and Asp-N work equally well. In general, as illustrated for TIM in Fig. 2, trypsin and Lys-C digests have several peaks in common but use of the latter enzyme results in fewer peptides; several of them, however, are late-eluting. Chymotrypsin treatment has a comparable result although no large fragments have ever been obtained. Subtilisin digests yield predominantly small peptides and are only recommended when everything else fails. We have insufficient experience with endoprotease Asp-N at this point to generalize its usefulnes, but tests on BSA and

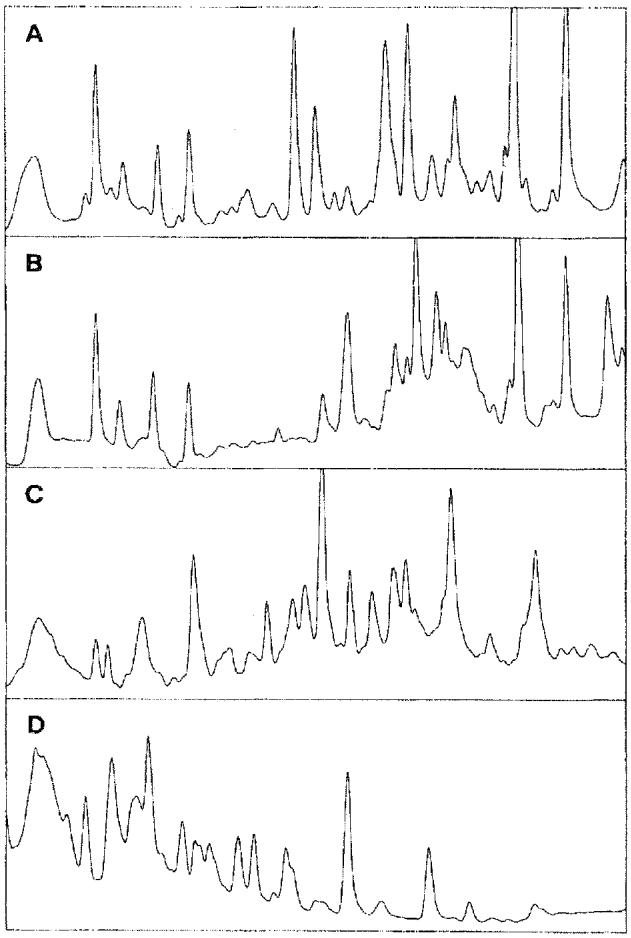

Figure 2. HPLC profiles of *in situ* digest of 150 picomoles triose phosphate isomerase, separated on a 1-DE gel and electroblotted on NC, in 25 µL of 0.1 M NH_4HCO_3/5 % acetonitrile with 1 µg of trypsin (panel A), endoproteinase lys-C (B), chymotrypsin (C) and subtilisin (D) for 15 h at 37 °C. HPLC was done using conditions listed in Section 2.9. Full scale is 0.05 AUFS; time scale is from 30 to 50 min.

TIM were promising and gave recoveries and patterns of complexity comparable to trypsin. In our hands, *in situ* digests using endoprotease Glu-C (*S. aureus* V8 protease) were unsuccessful.

The availability of several proteases for this technique improves the odds of recovering optimal size peptides (> 15 amino acids). Fragments of this length are preferred for oligonucleotide probe design for cloning. On those occasions where some 100 picomoles of substrate are available, parallel digests can be done to increase the number of sequenceable peptides. In the more frequent cases, when dealing with limited amounts of protein, we have successfully determined the protease of choice by doing pilot digests on low- to subnanogram quantities of NC-bound, ^{125}I-labeled substrates. Labeling was done, at an earlier stage of the partial purification of the protein, on a few percent of the total material. The results of two such experiments, on yeast 6-P-fructo-2-kinase (96 kDa) and inositol triphosphate receptor (230 kDa) from bovine aorta, are listed in Table 1. In both experiments, most counts were recovered from the NC strips after subtilisin digests (about

85 %) and the least counts with endoproteinase Lys-C (51 %). Trypsin and chymotrypsin treatments released between 65–70 % of the counts. The decision, in both cases, to use trypsin for a micropreparative digest was a compromise between better recovery (than with Lys-C) and the putative presence of larger fragments (than with subtilisin). Micropreparative experiments were spiked with radiolabeled proteins; recoveries from the NC were 75 % for the kinase and 60 % for the receptor protein.

Source and handling of the different proteases is crucial for the success of *in situ* digests. By testing many of the brands available in the US, we found drastic differences in quality, including some that had no activity at all. At present we prefer trypsin and *Pseudomonas fragi* endoproteinase Asp-N, both "sequencing grade" from Boehringer, and *Achromobacter lyticus* endoproteinase Lys-C (also known as lysyl endopeptidase) from Wako Chemicals. No real differences were observed between chymotrypsin and subtilisin from different suppliers. The long-term storage of enzymes is as lyophilized powders at 4 °C. Occasionally, small amounts are dissolved in 0.1 M NH_4HCO_3 in a 1 μg/μL concentration and divided into 5 μL aliquots that are stored frozen. The five enzymes listed above have been stored this way for up to a year without significant loss of activity. When needed, the solutions are thawed on ice for one-time use. Because of the high cost, only small quantities of endoproteinase Asp-N are purchased at a time, stored lyophilized and dissolved just before use. It is recommended to plan several simultaneous experiments when using this enzyme.

3.2.1.3 Digest reactions

Little is known about the catalytic processing of NC-bound proteins by proteases. Speculations have been made but, to our knowledge, no systematic study has ever been conducted on the structural changes of proteins immobilized on NC, the resulting substrate/protease interactions or the kinetics of the reaction. In general, the digests are kept at 37 °C for 15 h. A comparison of HPLC profiles resulting from two tryptic digests of 75 picomoles BSA, either in solution or electroblotted onto NC (after 1-DE), is presented in Fig. 3 (panels A and B). Although many identical peaks can be observed, the digest profile of the NC-bound protein contains fewer of them; some late-eluting, presumably larger peptides, are missing. Similar observations were made for digests with other proteases and substrates. Whether this is due to partial inaccessibility of the substrate or lowered activity of the enzymes is unknown. The presence of a cosolvent (acetonitrile) is an unlikely explanation for this observation for it has been shown that 5 % MeCN in bicarbonate buffer does not affect the activities of the proteases used in our study [37]. The properties of the eluting peptides and those that do not desorb from NC in 0.1 M NH_4HCO_3/5 % MeCN are also not precisely known. The longest peptide recovered in our laboratory was 25 residues in length. Of those that do diffuse out of the NC, peak heights are about 70–80 % of the corresponding peaks from equimolar amounts of the same substrates digested in solution (see Fig. 3, panels A and B).

The elution process can be quantitated by monitoring the release of counts in the solvent upon digestion of radiolabeled proteins. The results presented in Table 1 indicate that 50–85 % of the counts, depending on substrate and enzyme, are released. A comparable release of 65 % with α-lactalbumin (14.2 kDa) has been reported [20]. The results for the inositol triphosphate receptor were much better (55–85 % release) than we had expected for a membrane protein of such size (230 kDa); the resulting peptide pattern on HPLC was therefore complex (result not shown). Patterns are usually not too busy and single peaks correspond with single sequences. Occasionally, two or three peptides coelute within one symmetrical peak. Based on the presence of additional peaks over background, we have always observed release of peptides, although sometimes not quite as many as could be expected for a protein of a certain molecular weight. In all cases tested (TIM, BSA, carbonic anhydrase, ovalbumin and β-lactoglobulin), the profiles were reproducible.

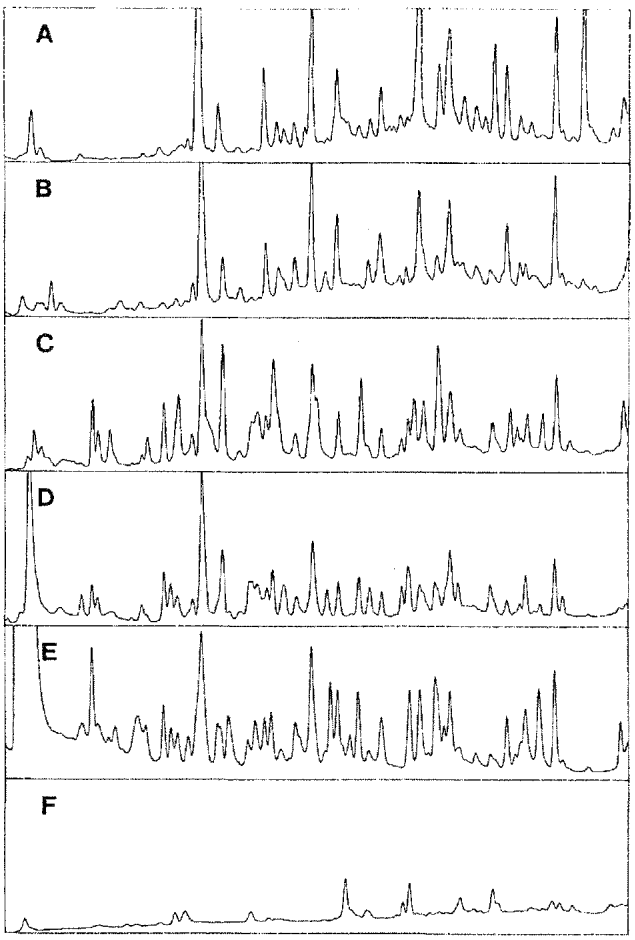

Figure 3. HPLC profiles of tryptic *in situ* digest of 75 picomoles BSA. All digests were done in 25 μL 0.1 M NH_4HCO_3/5 % MeCN and 1 μg trypsin. HPLC was done using conditions as described in Section 2.9. Full scale is 0.05 AUFS and time scale is 30 to 70 min. Panel (A) digest in solution. (B) *In situ* digest on NC after 1-DE separation. (C) BSA was reduced and alkylated in Laemmli buffer and immediately loaded on the gel, electrophoresed, electroblotted and digested as described in Section 2. (D) As (B) but the NC-immobilized BSA was reduced for 20 min with 0.5 % β-mercaptoethanol (37 °C) and 20 min alkylation with 1.5 % 4-vinylpyridine (25 °C/dark) in 250 mM Tris-HCl, pH 8.5, after the PVP-40 blocking step was done; reagent was removed by 5 washes with 5 % MeCN (in water) before the digest. (E) Reduction and alkylation was done in solution after the *in situ* digest as described. (F) Blank tryptic digest using 10 mm^2 of NC.

Table 1. *In situ* proteolysis of NC-bound proteins[a)]

Protein	Molecular mass kDa	Estimated Amount	Total CPM	Enzyme[b)]	After proteolysis %CPM NC	%CPM buffer
6-Phosphofructo-2-kinase (PF2K)	96	10ng	3500	–	48	5
		10ng	3500	T	15	65
		10ng	3500	KC	35	51
		10ng	3500	C	25	67
		10ng	3500	S	3	82
		2µg	15000	T	13	75
Inosital triphosphate receptor (IP3R)[c)]	230	0.5ng	9500	T	31	69
		0.5ng	9500	KC	44	56
		0.5ng	9500	C	23	77
		0.5ng	9500	S	14	86
		20µg	60000	T	40	60

a) The columns list (in order): protein and molecular mass; weight amounts and radioactive counts, associated with the protein, bound to the NC membrane; enzyme used for the digest; radioactive counts, after the digest, either still associated with the NC or released in the buffer. All digests were done for 15 h at 37 °C in 100 mM NH_4HCO_3/5 % MeCN with 1 µg of enzyme. Volumes were 10 µL for all 6-phosphofructo-2-kinase samples, 25 µL for the 0.5 ng amounts of inositol triphosphate receptor and 50 µL for 20 µg of the receptor. BSA (25 µg) was added as a carrier to 0.5 ng receptor during the digest.

b) T, trypsin; KC, endoproteinase lys-C; C, chymotrypsin; S, subtilisin

c) Bovine smooth muscle

3.2.2 Ultramicroscale digests

To be useful, the combined 2-DE/protein sequencing technique must allow analysis of all protein spots. Several practical obstructions will make this a difficult and tedious task (see Section 3.1). It is therefore important to increase the overall sensitivity of the technique to allow internal sequence analysis of only submicrogram amounts of protein on a gel.

3.2.2.1 Practical considerations

At the submicrogram level, we found it crucial to maximize protease concentration and to keep the amount of NC as limited as possible. Diluted enzymes (< 0.04 µg/µL), with insufficient amounts of protein substrate to act as a carrier (< 0.1 µg/µL), have a markedly reduced activity and an increased tendency to stick to the walls of the polypropylene Eppendorf tube. Moreover, PVP-40 blocking of NC is never 100 % (R. Aebersold, personal communication) so that enzymes are adsorbed out of solution proportionally with the amount of NC. In either case, the immobilized protease is unavailable for catalytic reaction. Increased amounts of NC will require larger volumes of buffer for submersion, diluting the protease even further. It is better to have a large amount of protein on one small NC strip than little protein on several pieces. Small solvent volumes will also prevent peptides from sticking to the reaction vial. There should be absolutely no drying steps between digest and sequencing. Ideally, the digest mixture is immediately injected for HPLC analysis, although storage for a few days at –70 °C does not seem to result in a significant loss of peptides.

Experiments where 2 nanogram quantities of radiolabeled 6-phosphofructo-2-kinase were digested with 1 µg of different proteases, in 10 µL of bicarbonate buffer/5 % MeCN, are given in Table 1. The NC was spun down and rinsed once with the same solvent; supernatant and washing solution were combined and the gamma-radiation measured immediately. About 65 % of the counts, associated with the NC strips, were recovered in the solvent in the absence of any carrier but the protease. The combined counts of NC and buffer did not add up to 100 %, indicating that losses on the tube walls had occurred. During an experiment with 0.5 ng inositol triphosphate receptor (1 µg protease; 25 µL volume), in the presence of 25 µg BSA as a carrier, all counts were present on the NC and in the buffer.

When working with an estimated 10–30 picomoles total immobilized protein (one spot/band, or several), the following practical guidelines have assured successful digests. Volumes are kept as low as possible and should never be more than 25 µL in a 500 µL Eppendorf tube. NC spots/bands are trimmed to the limit and should be sufficiently solvent-exposed and completely submerged; about 25 mm^2 of NC (cut in 2 mm^2 pieces) can be suspended in 25 µL (1 mm^2/µL) this way. The NC pieces are never allowed to dry; cutting and trimming are done with the NC submerged in water. Enzyme concentrations are always 0.04 µg/µL (1 µg/25 µL) or more. In the case of nanogram quantities of protein, the enzyme/substrate ratio is often bigger than 1.

3.2.2.2 Background considerations

A literature survey indicates that standard enzyme/substrate ratios are 1/20 to 1/100 and for NC-bound substrates about 1/5 to 1/20. There is a common worry that, when using more enzyme, there will be a background due to autodigestion of the protease that will totally obscure substrate-derived peptides. Indeed, with HPLC configurations and sensitivities permitting real time detection and collection of 5–10 picomole amounts of peptides (see Section 4.3), background peaks (resulting from 1 µg enzyme) and substrate-derived peaks are equal in size. This is clearly illustrated in Fig. 4 (panels A, B and C), where tryptic micropreparative digests of protein spots, from a 2-DE separation of total *E. coli* extract, are shown alongside a blank experiment using the same amount of NC (cut from a blot; no protein) and enzyme.

Although this is a cause for legitimate concern and could potentially lead to sequencing many protease-derived peptides, it is a problem that can be dealt with. We noticed that the background due to autoproteolysis of 1 μg enzyme is generally reproducible, with consistent elution times of all components and only slight changes in peak heights. When a parallel blank is done under exactly the same conditions as for the real digest, background peaks can be successfully eliminated by either visual or PC-based postrun processing of the chromatogram. Peaks for sequencing experiments are then chosen from the resulting pattern. By doing so, we have never sequenced an artifact peptide during analysis of several dozens of peaks from more than 10 microdigests.

Occasionally we have sequenced a peptide with a similar elution position as a background component. Reasons for these decisions were either a dramatic increase in peak height and/or the unquestionable additional presence of Trp, Tyr or *S*-β-(4-pyridylethyl)cystein (PE-Cys), as monitored with a diode array detector (see Section 3.4), over background. Since the sequences of trypsin [38], chymotrypsin [39], subtilisin [40] and *Achromobacter* Lys-C [41] are known, careful analysis of the sequencing result made it possible to differentiate the enzyme-derived background from the sequence of the unknown peptide. The analysis of peak 10 from the tryptic digest (Fig. 4, panel B) of a protein spot recovered from a 2-DE gel (Fig. 1, spot 1) is an exellent illustration of the feasibility of this approach. Despite the presence of a major trypsin-derived peptide (7 pm initial yield), a minor substrate-derived peptide (1.5 pm IY) was successfully sequenced and used to identify the protein. This experiment will be discussed in more detail in Section 3.6 (Applications). We were unable to retrieve the sequence of *Pseudomonas* proteinase Asp-N from the literature; this limits its practical use for these experiments at present. The current batch of Boehringer sequencing grade trypsin is not from a bovine source, as stated by the manufacturer; limited sequence analysis in our laboratory indicated a 100 % similarity to the pig trypsin primary structure.

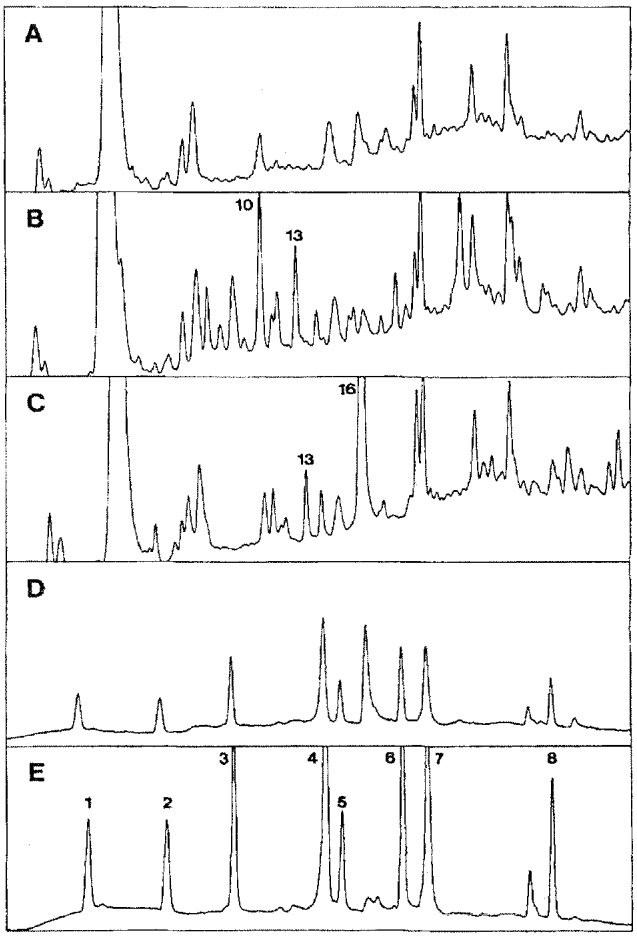

Figure 4. HPLC patterns of standard peptides and *in situ* digest of proteins separated on 2-DE gels. HPLC was done using conditions as described in Section 2.9. Full scale is 0.01 AUFS and time scale is 30 to 65 min. Panels (D) and (E) show separations of mixtures containing 5 and 10 picomoles, respectively, of the following peptides (except for peptide 5 that was present in 2.5 and 5 picomole quantities). (1) LKPTPE; (2) IFVQK; (3) TGQAPGFTYTDANK; (4) YIPQPRPPHPRL; (5) peCPSPKTPVNENNPQ; (6) ILLQKWE; (7) YSLEPSSPSHWGQLPTPP-NH2; (8) GITWKEETLMEYLENPK. Panels (A)–(C) are HPLC profiles of *in situ* digest in 20 μL 0.1 M NH_4HCO_3 with 1 μg trypsin for 15 h at 37 °C. (A) Blank NC; (B) spot 2d≠1 (Fig. 1) from a Ponceau S-stained NC electroblot of a 2-DE separation of 120 μg total protein from *E. coli*; (C) as (B), but spot 2D≠2. Peaks 10 and 13 (panel B) and 13 (panel C) were sequenced.

3.3 Cysteine derivatization

There is a popular belief that after SDS-PAGE under reducing conditions, proteins are completely unfolded and all disulfide bridges broken. We have reason to believe that this may not always be the case (see Section 3.2.1.1) and could result in incomplete proteolysis and the recovery of disulfide-linked peptides after HPLC: poor yields and mixed sequences are the dire consequences. In addition, unmodified cysteines are chemically unstable and cannot be sequenced well; no positive identification is possible at the low picomole level. The resulting gap in the determined sequence has had disastrous consequences for probe design on many occasions. A selective, high-yield, alkylating reagent, in combination with a reductant, should be used to derivatize the reactive sulfhydryl groups. 4-vinylpyridine is often used to this purpose [42, 43]. This reagent converts reduced cysteines to their stable *S*-pyridylethyl derivatives that are compatible with Edman chemistry [43] and easily identified. The derivative has a strong absorption maximum at 253 nm which allows tracking of PE-Cys-containing peptides during liquid chromatographic separation ([44]; Fleming and Tempst, unpublished results). Reduction and *S*-alkylation of proteins with 4-vinylpyridine requires removal of excess reagent and byproducts after reaction. We have observed undesirable side reactions and PE-Cys degradation, even at −20 °C, in cases where this was not done immediately. Dialysis or RP-HPLC are simple desalting procedures suitable to this purpose. However, the losses of low picomole amounts of proteins, due to unspecific sticking to dialysis bag or column, are unacceptably high. Alternatively, the alkylation reaction has been done *in situ* on glass fiber filters where the reagent is either delivered as a vapor [45] or efficiently extracted with organic solvent [43].

With all the above considerations in mind, we set out to determine a strategy and optimal conditions for efficient reduction and *S*-pyridylethylation of proteins separated on 1-DE and 2-DE gels and destined for internal sequence analysis. BSA (67 kDa) is particularly well suited for this study since it cotains 35 cysteines, involved in 17 disulfide bonds [46]. In principle, the

reaction can be carried out either immediately before the gel is run, or *in situ* on NC, or in solution after the digest just prior to HPLC. Patterns of tryptic digests on 75 picomoles BSA, pyridylethylated at these three various steps in the total protocol, are compared in Fig. 3 (panels C, D and E). Profiles of digestions on equal amounts of unreacted BSA, either in its native form in solution (panel A) or electroblotted onto NC after SDS-PAGE under reducing conditions, are shown for comparison. A tryptic digest on blank NC was also done (panel F). One μg of trypsin was used in 25 μL solvent for all digestions; other specific experimental conditions are given in the figure legend and in Section 2 (Methods).

Analysis of the results in Fig. 3 leads to two major observations. Digests of reduced and alkylated BSA invariably yield more peptide peaks (34–36) as compared to the standard *in situ* digest (about 25 peaks). Spectral analysis of all 36 peptides indicated that 22 contained at least one PE-Cys residue (results not shown). There is no direct evidence whether the increase in number of peaks is caused by the cleavage of remaining disulfide bonds, thereby separating covalently linked peptides, or by incomplete derivatization of single cysteine-containing peptides, resulting in two different species of each. However, double reaction times for reduction and alkylation did not result in different peak patterns. Additional tests with cysteine-containing synthetic peptides in solution, under exactly the same conditions as used during pre-gel or post-digest alkylations, indicated that the amounts of reagents were adequate to ensure a nearly 100 % derivatization. Thus, leftover disulfide bonds are the most likely explanation for our observations. Whether this is due to incomplete reduction or whether bonds are reformed after blotting, staining and destaining, is not known.

Judging from the lowered peak heights during HPLC (Fig. 3, panel D) the *in situ* alkylation procedure resulted in a serious decrease of yield. This is most likely due to protein washout during the 40 min incubation at pH 8.5. Protein losses from NC have previously been associated with alkaline pH [20]. Final yields of peptides, generated using the procedures that contain either a pre-gel (panel C) or post-digest (panel E) *S*-alkylation step, were similar to those resulting from an experiment where derivatization was omitted (panel B). Recovery and efficiency are the same for both alkylation procedures. With post-digest alkylation, a 15–20 min wash step (isocratic at 5 % solvent B) is required after loading the column and before starting the gradient. This effectively removes β-mercaptoethanol and 4-vinylpyridine that cause terrible baseline disturbances and obscure peaks. Actually, we found this isocratic step to be advantageous for the separation of *in situ* microdigests in general; it has become standard procedure. In conclusion, pre-gel *S*-alkylation is a simple and efficient method to facilitate internal sequence analysis, with maximal information, of proteins separated on 1-DE gels. We prefer post-digest derivatization for proteins recovered from 2-DE gels.

3.4 Liquid chromatography

For the liquid chromatographic separation of peptide mixtures at the picomole level, sensitivity is a primary consideration in selecting an instrument and the use of narrow-bore (1–2.1 mm inner diameter) columns is a prerequisite. Practical aspects of narrow-bore HPLC of peptides have recently been reviewed in detail [47, 49]. Based on the authors' conclusions and our specific requirements, we assembled a modular system as described in Section 2. Our aim was to evaluate and optimize performance at the 5–10 picomole level. For reasons that will be discussed in Section 3.5, it would be a significant advantage to have knowledge of the presence of PE-Cys and Trp in a peptide prior to sequencing, especially at the 2–5 picomole level. Diode-array detection has been used successfully for identification of Trp- or Tyr-containing peptides [49, 50]. Here, we describe our experience and practical hints for narrow-bore HPLC, in combination with diode-array detection, for purification and analysis of low picomole (<20) quantities of peptides.

3.4.1 Narrow-bore RP-HPLC

Originally all experiments were done using an ABI 783A UV-detector. When this detector was replaced with an ABI 1000S diode-array detector, sensitivity at 214 nm and resolution were equivalent (Fleming and Tempst, unpublished observations). We used a mixture of 8 synthetic peptides, varying in length from 5–18 amino acids, to study the sensitivity of the system. The lengths were specifically chosen to resemble the sizes of peptides typically released during *in situ* digests. Decreasing quantities of this peptide mixture were chromatographed under similar conditions. Peak heights were proportional to injected amounts, from 50 down to 10 picomoles, except for peak 8 (Fig. 4, panel E) which decreased about twofold more, relative to the others. However, when 5 picomoles were injected, all peak heights were only 70 % of the predicted values as based on the calibration curves derived from the previous experiments. This result is reproducible and a serious reason for concern. We do not know the underlying reasons for it, but unspecific sticking to the column is a major possibility. Studies are under way in our laboratory with even smaller quantities of peptides and with varied column supports and lengths, gradient slopes and injection volumes. Standard chromatograms for 5 and 10 picomoles of 8 peptides are shown in Fig. 4 (panels D and E). Despite the mentioned problem, at the 5 picomole level all peaks are well above baseline noise and it is conceivable that, by postrun computer enhancement of the digitized signals, subpicomole quantities could be detected. In fact, this has been reported at the analytical level [51]. For micropreparative experiments, however, the signal must be detectable in real time to permit manual collection of the peptide.

Appearance of peaks is followed on a stripchart recorder with its full scale corresponding to 0.01 or 0.02 absorption units at 214 nm. At these sensitivities, great care must be taken to ensure a flat baseline; 10–20 μL aliquots of neat TFA are added to 1 L of solvent A or B until this is the case. At a flow of 100 μL min, peaks will typically elute in 40–60 μL of solvent. This volume is about equal to that of a drop forming at the end of the outlet tubing. When the beginning of a peak is observed, the forming droplet is quickly removed with a kimwipe and the fraction collected by holding the end of the tubing, so that it just touches the wall of the micro-Eppendorf tube. The acetonitrile-containing solvent has a low viscosity and flows easily to the tip of the tube. No droplet is formed; this limits the collection volume and allows efficient collection of closely eluting peaks. Of course, the delay time between flow cell and collection should be minimal. We therefore plumbed an uninterrupted, stainless-steel piece of outlet tubing (18 cm; 0.005

inch inner diameter) directly into the flow cell. The lag time, with a flow of 100 μL/min, is 1.4 s. The narrow-bore tubing also ensures adequate back pressure in the flow cell to prevent gas bubble formation; samples do not need degassing. Collected fractions are immediately put on ice and frozen on dry ice after the run.

When both operated under optimal conditions, 4.6 mm ID columns give better resolution than their 2.1 mm counterparts [47, 49]. Unfortunately, only the latter type columns are useful in our research. A limited investigation indicated that this problem will not be cured easily without improvements of the stationary phases and column packing techniques. We unsuccessfully attempted to improve resolution by varying gradient slopes, temperature, loads and injection volumes and columns. At present, only two-dimensional RP-HPLC, using different columns or solvent systems, can provide the required resolution of complex peptide mixtures. We found that, due to some practical problems, at least 20–25 picomoles of peptide are needed for this technique. For smaller quantities and in the absence of a better column, we prefer a single chromatographic run using an Aquapore RP-300 support. The reproducibility with this column and system is excellent and allows postrun background subtraction, needed before selecting a peak for sequencing (see Section 3.2.2). An example can be found in Fig. 4, where tryptic digests are shown (panels B, C) alongside an enzyme blank (panel A); background peaks are easily spotted.

3.4.2 Diode-array detection

Tryptophan has a unique codon and its presence in a peptide sequence facilitates construction of oligonucleotide probes with low sequence degeneracy. Gene cloning strategies, involving the use of such probes, have become popular among molecular biologists during recent years [1]. Trp-containing peptides can be screened for by dual wavelength monitoring of HPLC eluates at 214 and 280 nm. However, there is interference from other aromatic amino acids, such as Tyr, and from chemically derivatized residues such as *S*-pyridylethyl Cys. Diode-array detection and the use of second-order derivative spectroscopy allows easy identification of Trp or Tyr [49, 50]. However, we could not unambiguously differentiate between these two amino acids at the 5–20 picomole level. In addition, with complex peak patterns, routine analysis of all peaks becomes very time consuming.

We are currently using a simple method that permits fast screening of all peaks. The column effluents are monitored simultaneously at 214, 253, 277 and 297 nm and the peak heights compared. We have determined simple empirical rules, using these values, that allow unambiguous identification of Trp-, Tyr- and PECys-containing peptides at the 5 picomole level (Tempst, Fleming and Lane, unpublished results). Examples, using synthetic peptides, are shown in Table B. Relative absorptions of Pe-Cys at the different wavelengths are as follows: A253>A277>A297; this series is A277>A253>A297 for Trp and Tyr. The latter two can be easily distinguished because the A297/A277 ratio is about 25–30 % for Trp and less than 3 % for Tyr. This comes down to no measurable absorption, at 297 nm, for Tyr-containing peptides below the 50 picomole level. Absorption values can be obtained from the spectra (with A320 as the arbitrary zero value) or, more easily, by measuring peak heights on a stripchart recording. We found that this can be done accurately for peptide amounts as low as 10 picomoles (see Table 2). Peaks cannot be measured at the 5 picomole level but an observed deflection of the baseline at 253 or 297 nm allows positive identification of PE-Cys or Trp, respectively. A more detailed account of these studies will be published elsewhere (Tempst *et al.*, in preparation).

In practice, a quick visual scan of the four chromatograms allows immediate identification of Trp-, Tyr- or PE-Cys-containing peptides; spectra of these peptides can then be inspected in detail if desired. As shown in Table 3, eight predicted tyrosines, at the 4–10 picomole level, were all confirmed by sequencing. Only two out of the five predicted tryptophans were identified during sequencing at these low levels. A third Trp, present in peptide inositol triphosphate receptor 48.2, was later confirmed when an identical sequence was retrieved from the National Biomedical Research Foundation database.

When *in situ* digests of NC-bound proteins are chromatographed, especially when reduced and alkylated just prior to HPLC, a 15 min isocratic elution step (at 5 % B) before starting the gradient, is recommended. It will remove most non-peptide contaminants and ensure flat baselines at all 4 wavelengths. Occasionally, with sensitivities at 0.001 absorption units full scale (AUFS), ghost peaks appear. An example can be seen in Fig. 4 (panel D) between peaks 5 and 6 and at the corresponding position in panel C. Analysis indicated that these peaks did not contain any polypeptide material.

Table 2. RP-HPLC with multiple wavelength detection of peptides containing Trp, Tyr or Cys[a)]

Peptide	Pico-moles	Column ID mm	Peak height (mm) A_{253}	A_{277}	A_{297}	W	Y	PEC
P89-05PE	800	4.6	244	28	6			+
P88-60	800	4.6	122	280	84	+		
P88-53	800	4.6	38	116	3		+	
P88-54PE	30	2.1	88	88	18	+	+	+
Mix≠4	10	2.1	2	10	0		+	
Mix≠5	10	2.1	12	2	0			+
Mix≠6	10	2.1	15	35	9	+		

a) Peptides were chromatographed on a Vydac C4 (4.6×250 mm) or an Aquapore RP300 (2.1×220 mm) column at flows of 1 and 0.1 mL/min, respectively. The HPLC system used was as described in Section 2.9. Peak heights on chromatograms, produced by monitoring at different wavelengths, are expressed in mm on a 25 cm recording with sensitivity set at 0.01 AUFS. The (+) under W, Y or PE-Cys indicated that this residue was present in the peptide. Peptide sequences were: P89-05PE, peCPSPKTPVNFNNFQ; P88-60, GNLWATGHF; P88-53, YEVKMDAEF; P88-54PE ISpeCWAQIGKEPITFEHINYERVSDR. Peptides Mix≠4, 5 and 6 are listed in the legend to Fig. 4.

Table 3. Spectral properties and sequencing results of peptides obtained by *in situ* digests on NC and HPLC separation[a)]

Protein	Molecular mass kDa	Peak≠	Peak weight (mm) A_{253}	A_{277}	A_{297}	Predicted W	Y	IY pmoles	Residues	sequenced	Remarks
6-Phosphofructo-2-kinase (PF2K)	96	5	ND					2.0	7/7		Gene cloned
		11	ND					1.4	14/16		
	93	13	ND					1.8	10/11		Gene cloned
		20	ND					3.1	15/15		
Phage SP6 DNA polymerase (SP6DP)	100	30	14	30	6	+		2.3	9/9	(-)	Gene cloned
		42	22	54	7	+	+	2.0	15/18	($Y_9Y_{10}Y_{13}Y_{15}$)	
Bovine brain protein (BD-43)	43	11	0	0	0			14.0	18/19	(-)	Identified
		14	58	148	26	+	+	9.6	12/12	(Y_1W_3)	
Inositol triphosphate receptor (IP3R)	230	48.2	46	110	14	+	+	3.0	13/14	(Y_{14})	Matched
Calcium activated serine protein (brain)	65	10	3	12	0		+	3.5	10/10	(Y_3)	
E. coli[b)] 2D ≠ 1	50	10	8	16	4	+		1.6	10/11	(W_8)	Identified
		13	0	4	0		+	1.5	12/14	(Y_{11})	
2D ≠ 2	40	13	1	7	0		+	1.0	5/9	(Y_1Y_9)	
2D ≠ 3	40	17	2	10	0		+	1.7	10/11	($Y_1Y_8Y_9$)	Identified
		18	3	12	0		+	1.0	12/15	($Y_1Y_7Y_{12}$)	

a) Peak heights (in mm at 0.01 AUFS) at different wavelengths are listed and the presence of Trp (W) or Tyr (Y) predicted. IY, initial coupling yields during sequencing. Residues identified/total cycles analyzed with identified Tyr or Trp between brackets. The last column lists whether the protein was identified or cloned: BD-43 is glutamine synthetase; *E. coli* 2D≠1 is elongation factor Tu; *E. coli* 2D≠3 is an outer membrane porin protein.

b) *E. coli* 2D≠1 to ≠3, spots 1–3 on a 2-DE gel of total *E. coli* protein (see Fig. 1). Peak ≠ indicates the peak collected after *in situ* digest and HPLC (shown for *E. coli* 2D≠1 and ≠2 in Fig. 4, panels B and C).

3.5 Polypeptide sequencing

Narrow-bore RP-HPLC allows recovery of 4–6 picomoles of peptides which can be used for direct sequencing. The initial coupling yields during analysis could then be as low as 1.5–2 picomoles and, due to washout and a slight inefficiency of the Edman reaction, the signal will quickly drop below the picomole level. The analytical end-point method must therefore allow subpicomole PTH-amino acid identification if the analysis is to be continued for another 10 or more cycles. In general, the method of choice is on-line narrow-bore HPLC [2]. Rare examples have appeared in the literature where subpicomole sequencing was done; none described how the sequence calls were exactly made. It is fair to state that, until recently, sequencing with signals at the subpicomole level was extremely difficult. We have recently introduced simple modifications of existing automated analysis techniques to improve the results obtained during subpicomole sequencing and to make the process more routine. This resulted in an eightfold increase in sensitivity over standard methods. Details of this investigation are beyond the scope of this report and can be found elsewhere [3]. A reference chromatogram for 700 femtomoles PTH standard, produced in real time on a stripchart recording, is shown in Fig. 5.

Several low picomole quantities of peptides, from *in situ* digests of proteins separated on polyacrylamide gels, have been sequenced in our laboratory. Selected examples are listed in Table 3. During most of these experiments, PTH-amino acid signals dropped below the picomole level at some point; analysis was continued until no further signals were observed. This yielded 10–15 amino acids of sequence in most cases. A few peptides were smaller and were sequenced completely (as judged from a C-terminal Lys or Arg for tryptic peptides). Of

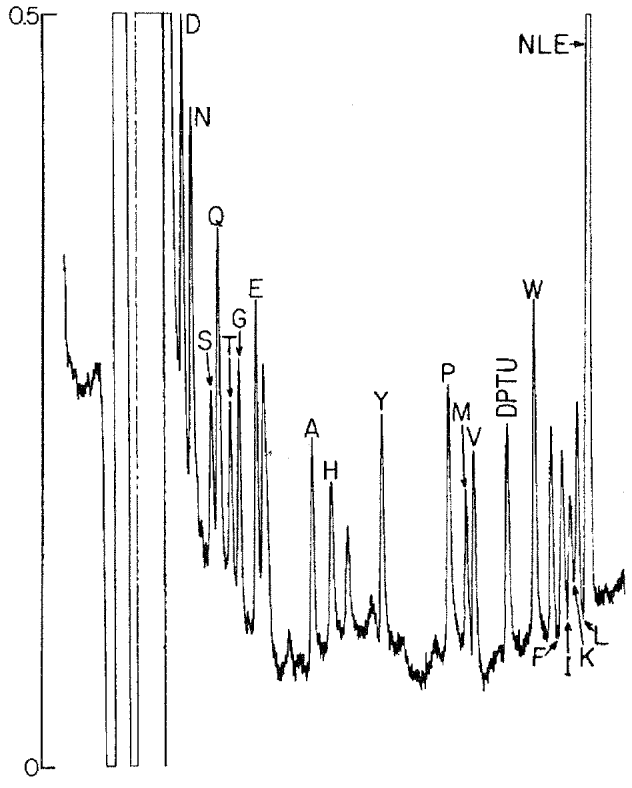

Figure 5. RP-HPLC analysis of 700 femtomoles PTH-amino acid standards using conditions described in Section 2.9. Full scale corresponds to 0.0005 AUFS. Reprinted from [3], with permission.

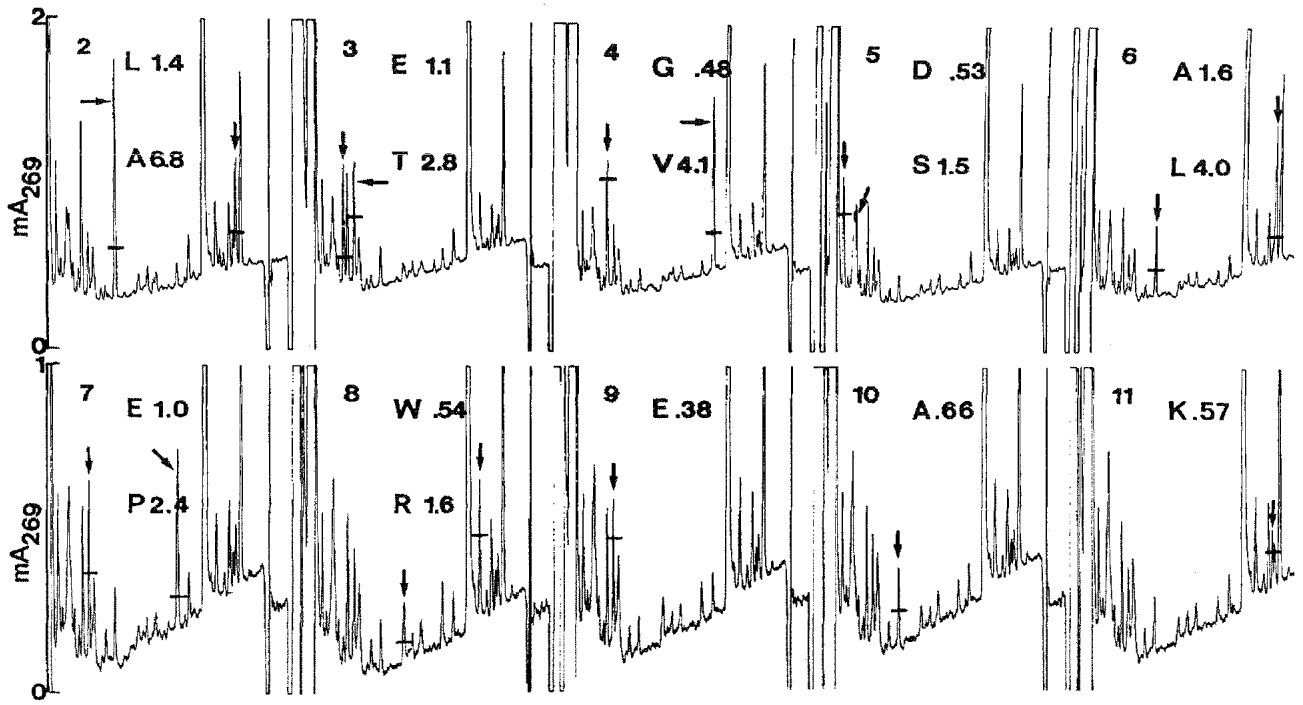

Figure 6. Amino acid sequence analysis of peptide 10 (Fig. 3) from an *in situ* digest of spot 2D≠1 from a 2-DE gel separation of total *E. coli* proteins (Fig. 1, ≠1). Conditions were as listed in Section 2. The amount of peptide for analysis was unknown. Chromatograms (2) to (11) are shown; full scale corresponds to 0.002 AUFS for cycles (2)–(6) and 0.001 AUFS for cycles (7)–(11). Increases of peaks, compared to the previous cycle, are indicated (portion above the horizontal bar) and the picomolar quantities are shown. PTH-amino acid values are background-subtracted. A mixture of two peptides was present: the sequence listed at the top is from the *E. coli* protein and is identified to be elongation factor Tu; the one at the bottom is a peptide from porcine trypsin.

particular interest are two sequencing experiments on peptides obtained after digestion and chromatography (Fig. 4, panel B, peaks 10 and 13) of a single spot from a 2-DE gel separation (Fig. 1, spot 1). Peak 10 coelutes with a background peak (see panel A); not surprisingly, analysis yielded two sequences (Fig. 6). The major sequence of the mixture matched the porcine trypsinogen primary structure (positions 108–115). This allowed easy determination of the minor sequence. When sequencing peak 13 (Fig. 7), all calls were made at the subpicomole level, except for Ile at position 3 (1.1 picomole; valu used to calculate IY) and for cycles 1 and 2 where no identification could be made due to excessive background. Initial yields for both sequencing experiments were around 1.5 picomole. Judging from the peak heights during HPLC (Fig. 4), relative to those of 5 and 10 picomoles of standard peptides, about 5–6 picomoles were collected. Initial yields were therefore 25–30 %, not taking into account possible losses during collection, storage and handling. Based on our experiences, peptide peaks from the mid-part of the chromatograms, with absorptions > 0.004 OD (at 214 nm), are good candidates for reliable sequencing. We observed that, although no special precautions had been taken during electrophoresis and electroblotting, no modifications or destruction of Trp, Met or other amino acids had occurred, not even at the low-picomole level.

Routine operation of the PTH-amino acid analyzer, for all applications discussed in this report, was at a sensitivity of 0.002 AUFS; this is quite often changed to 0.001 AUFS in the course of an experiment, when judged necessary (*e. g.* during the runs shown in Figs. 6, 7). Artifactual peaks, due to contaminatinating amino acids and by-products of the Edman chemistry, constitute a major problem for analysis at this level. Amino acid background usually prevents unambiguous calls during the first few cycles. Chemical background can be limited by changing reagents and solvents often and by omitting DTT from solvents S1, S2 and S3. The machine is thoroughly cleared 3 to 4 times a year. We found it beneficial to set the R2 regulator at 0.3 psi. This will decrease the unwanted delivery of water vapors to the reaction cartridge during the coupling step and results in drastically lowered diphenylthiourea and diphenylurea peaks. It permits positive identification of Trp at the subpicomole level (see Fig. 6, cycle 8). The associated reduction in repetitive yields is not a real disadvantage for sequence analysis of peptides smaller than 20 residues.

The position of PTH-PE-Cys in the chromatogram is dependent on the ionic strength. In our system, it is positioned around Pro/Met/Val. We have observed drifting of this peak, with aging of solvent A, resulting in coelution with one of the aforementioned other analytes. It is convenient to sequence PE-Cys-containing peptides first, after making fresh solvents, to avoid frequent ionic strength adjustments later.

3.6 Applications

Using the improved methods described in Sections 3.2–3.5, a number of proteins separated on 1-DE or 2-DE gels were

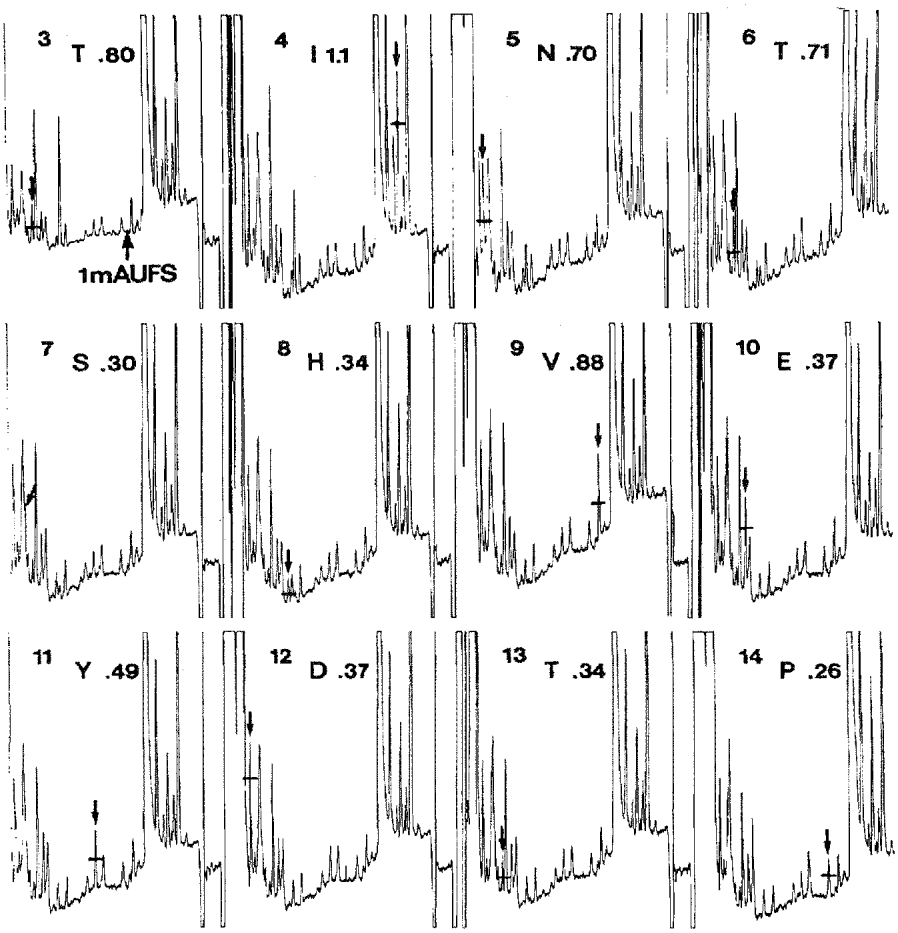

Figure 7. Amino acid analysis of peptide 13 (Fig. 4B) from an *in situ* digest of spot 2D≠1 from a 2-DE gel separation of total *E. coli* protein (Fig. 1, ≠1). Chromatograms (3)–(14) are shown; full scale was 0.002 AUFS but was changed during cycle 3 to 0.001 AUFS (arrow). For other details see legend to Fig. 6.

analyzed for internal sequences (Table 3). Except for protein BD-43 and the inositol triphosphate receptor, only submicrogram to less than 3 μg amounts of protein were present on the NC membrane before the digest. One or more peptides from each protein were successfully sequenced. The results allowed identification or cloning of the genes. Two species of yeast 6-phospho-fructo-2-kinase (93 and 96 kDa), previously found to be *N*-terminally blocked (Kretschmer, unpublished observation), were purified on a 1-DE gel, electroblotted and digested *in situ*. Oligonucleotides, derived from amino acid sequences of the two proteins, were used to isolate a single clone from a yeast genomic library. Since all probes hybridized to the same gene, the major 93 kDa species is probably a processed form of the 96 kDa protein (Kretschmer *et al.* in preparation). Similarly, internal sequences of electroblotted DNA polymerase (100 kDa) from phage SP6 allowed construction of two probes and quick identification and isolation of a 6 kb fragment from the phage genome (Rush *et al.*, unpublished results).

Impure preparations of protein BD-43 from bovine brain, isolated because of its presumed association with some other proteins (Denker *et al.*, unpublished observations), were separated on a 1-DE gel and analyzed. The NBRF protein database was searched with sequences of two peptides. They were found to be identical with stretches of the glutamine synthetase primary structure.

In situ digest of 20 μg NC-bound, bovine smooth muscle inositol triphosphate receptor yielded an extremely complex HPLC pattern and repurification of the peaks was necessary. Still, several peaks proved to be peptide mixtures. The first unambiguous single sequence that was obtained (peptide 48.2: 13 out of 14 amino acids; gap at position 12) was used for a database search and matched perfectly with the sequence of a partially characterized membrane phosphoprotein from mouse brain (Marks *et al.*, unpublished observation). The presence of a Trp in peptide 48.2 (Table 3) had been predicted, based on its spectral properties, but was not confirmed during sequencing. Interestingly, the gap at position 12 corresponds to a Trp in the known sequence of the brain protein.

3.6.1 2-DE Gels of *E. coli* total protein

An experiment was done to test the limits of sensitivity of the technique in an unbiased way. Total protein extract (120 μg) of *E. coli* was separated on a single 2-DE gel, electroblotted onto NC and stained with Ponceau S. Three out of the four major spots (indicated as 1,2 and 3 in Fig. 1; molecular mass 40–50 kDa) were arbitrarily chosen for internal sequence analysis. The amounts of each protein spot were estimated to be 0.5–1 μg (10–20 picomoles). Several peptide peaks, collected during separation of each digest (shown for proteins 2D≠1 and 2D≠3 in Fig. 4, panels B and C), were sequenced (sum-

marized in Table 3). Initial yields were between 1.0 and 1.7 picomoles and, with one exception, stretches of 10 or more amino acids were identified in all cases. The detailed sequence results of peaks 10 and 13 of spot 2D≠1 are presented in Fig. 6 and 7.

When all sequences were used to screen the NBRF protein database, we retrieved, to our pleasant surprise, two proteins of known structure, allowing identification of spot Eco2D≠1 as elongation factor Tu (tufA and B genes) and of spot Eco2D≠3 as a member of the outer membrane porin protein family from *E. coli* and phage lambda (ompF, 1c, nmpC and phoE genes). Incidentally, elongation factor Tu has an acetylated *N*-terminus [52] which would have prevented direct sequencing. The sequences of all peptides derived from proteins 2D≠1 and 2D≠3 were of such quality that they would have allowed the construction of low stringency probes of at least 20 nucleotides in length. The limited sequence of a peptide from protein 2D≠2 did not allow the spot to be identified and would have been unsuitable for probe design.

4 Gene cloning strategies

Two or more stretches of acquired internal sequence facilitate cloning of the gene. Depending on the size and abundance of the mRNA, at least three different strategies can be followed. Although quite often not sufficient by itself to allow cloning of the gene, the availability of amino terminal sequence may simplify cloning experiments and improve the odds for obtaining full-length cDNA and genomic clones. Despite the disappointing success rate, every possible effort should be made to obtain *N*-terminal sequence information. Cloning strategies are schematically presented in Fig. 8 and outlined below. Unless expert techniques are used, oligo-dT primed cDNA clones of long messages (>6kb) are quite often not full-length, seriously limiting the use of probes based on *N*-terminal sequence (Nt probes; Fig. 8.1 and [53]). A set of "internal" probes will improve the odds for isolation of clones and guarantee quick identification of "true positives". In general, use of multiple probes, including an Nt probe if available, will allow selection of clones spanning the longest stretch of coding sequences. Using a primer derived from the *C*-terminal part of an internal peptide, specific primed cDNA libraries can be constructed. Screening is done with either an Nt probe or internal probes derived from the amino-terminal part of the same peptide [54] or from different peptides (Fig. 8.2).

Finally, partial clones can be obtained by PCR amplification of first-strand cDNA using sense and antisense primers derived from two stretches of amino acid sequence (Fig. 8.3 and [24]. In the absence of another probe, fragments can be detected on the gel by hybridization with a labeled primer or by ethidium bromide staining. Having an *N*-terminal sequence is a preferable situation; PCR reactions with primers derived from all internal peptides can then be done and the longest fragment selected. Alternatively, primers can be derived from the *N*-terminal and *C*-terminal part of a single peptide [24]. A third possibility would be to derive both a sense and antisense primer from each peptide, and, since the relative positions of the internal peptides are not known, use them in all possible combinations. Fast oligonucleotide synthesis and parallel PCR reactions should easily accomodate this technique.

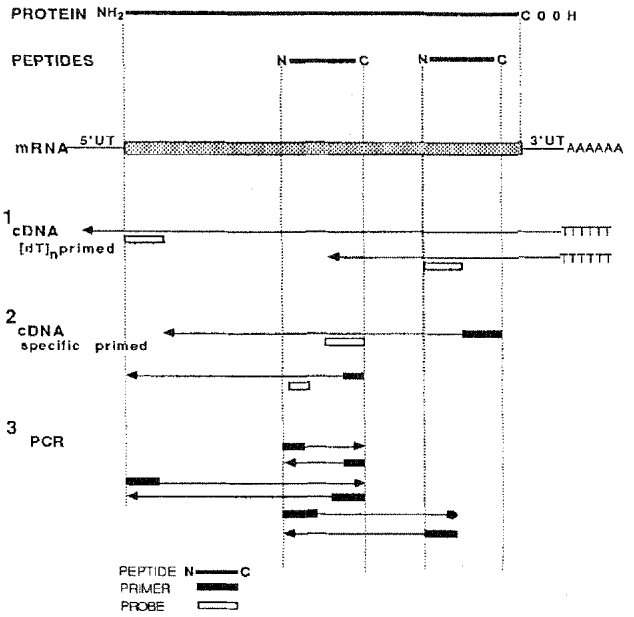

Figure 8. Cloning strategies using partial amino acid sequences. (1) Oligo-dT primed cDNA synthesis and screening with probes derived from *N*-terminal and internal amido acid sequence. (2) Specific primed cDNA synthesis and screening with internal probes derived from the same (or different) peptide that served to design the primer. (3) PCR application using first-strand cDNA and primers derived from amino terminal and different internal sequences.

5 Conclusions and perspectives

5.1 New trends in biology

During recent years, three interesting technologies have emerged in different fields of the biological sciences. Partial amino acid sequence information of proteins has increasingly been used to clone the corresponding genes through the use of synthetic oligonucleotide probes. Variant or changed cellular phenotypes have been examined, at the level of the single protein, using high resolution 2-DE and data for affected proteins stored in comprehensive databases. Finally, techniques were introduced for efficient sequence analysis of proteins separated on polyacrylamide gels. Theoretically, the integration of these three technologies would allow detection of phenotypical differences, partial amino acid sequence analysis of regulated or variant proteins, cloning of their genes and further studies of genomic variability and transcriptional regulation using DNA probes. For this approach to become a general research tool, one must be able to cope with all proteins, including those of lowest abundance. In addition, stretches of partial amino acid sequence should be uninterrupted and sufficiently long to allow construction of oligonucleotide probes. Unfortunately, one of the limiting factors is the efficiency and sensitivity of protein purification and sequencing.

5.2 Improved methods

In this report, we have described improved methods that allow extensive internal sequencing of 10–20 picomoles of protein recovered from a 1-DE or 2-DE gel. This means that, for most

proteins, only submicrogram quantities are required to start the analysis. Optimizations of *in situ* micro-digests and liquid chromatography were found necessary for successful handling and recovery of such small amounts. In addition, two new techniques were presented. First, a simple method was described for real time identification during HPLC of peptides containing Trp or derivatized Cys. Due to its unique codon, peptides containing Trp are preferred for design of low-degeneracy probes. Second, we found that *S*-alkylation of gel-purified proteins promoted efficient cleavage, eliminated sequencing of disulfide-linked mixed peptides and allowed positive identification of Cys. Automated protein sequencing with subpicomole analysis, which had been developed previously [3], was used.

5.3 Applications

Using our improved methods, a variety of unknown proteins, only available in minute quantities and purified on 1-DE gels, were sequenced, which enabled either positive identification or cloning of the genes. Internal sequences were also obtained from submicrogram amounts of two proteins, recovered after a single 2-DE gel separation of 120 µg total protein from *E. coli*, and allowed identification of the two selected spots as elongation factor Tu and outer membrane porin protein. Analysis of 90–95 % of the other spots will require specific enrichment before running the gels. Compared to a single amino terminal sequence, multiple stretches of internal sequence allow a more accurate identification of the proteins and offer clear advantages for conventional and PCR-based gene cloning. However, the availability of both internal and amino terminal sequences is most desirable. In general, the success rate of internal sequence analysis was satisfactory, but given the current minimal requirements (10–20 picomoles), only a limited number of proteins can be analyzed after 2-DE separation of total cellular extracts. Until femtomole quantities of polypeptides can be routinely recovered and sequenced, cell fractionation and column chromatography will be required to enrich most proteins of interest before gel electrophoresis. Two types of 2-DE experiments will therefore be needed: high sensitivity analytical gels with autoradiography or enhanced gold stain detection for phenotype analysis and micropreparative gels of partially purified fractions.

5.4 Future developments

The need for further improvements of all integrated technologies is apparent. Automated 2-DE will allow faster sample turnover with higher reproducibility. Hardware modifications and miniaturization of liquid chromatography instruments and automated sequencers are required for femtomole operation. Although chemical protein sequencing is still the most sensitive method today, the limits may have been reached [3]. Alternative technologies such as mass spectrometry (MS), while already being faster, will certainly become more sensitive and feature a prominent role in polypeptide sequencing during the next decade [56]. Fragmentation of the protein, similar to the method described in this report, is required for MS-aided sequencing. In general, cleaner environments will be needed for protein microseparations. A predictable aspect of increased technological sophistication will be the skyrocketing costs of hardware purchase and maintenance. As a result, execution of the entire procedure outlined here will become more restricted to specialized core facilities, if not already the case. In view of the resulting paucity of related research in other academic laboratories, research and development in the core laboratories must be emphasized and should coexist on an equal basis with analytical services. Independent funding will therefore be necessary. Graduate students and especially postdoctoral trainees should be encouraged and supported to participate in these developments and apply the emerging technologies and methods to biologically relevant problems.

5.5 Human genome sequencing project

Predictably, sequencing of proteins from 2-DE gels will quite often be done during case-by-case analyses of different cellular biological problems. On the other hand, systematic analysis of all protein spots of which quantitative and descriptive data are available in comprehensive databases could lead to positive identification and partial cloning of the genes. In either case, many DNA probes and markers, linked to some regulatory event or genetic condition, will be generated. A future integration of the 2-DE / sequencing / cloning studies with the ongoing human genome sequencing project [55] is therefore apparent and to be expected. The conversion of all 2-DE gel spots into sequence data for generating sequence-tagged sites (STSs) [55] will provide entry points into the human genome for constructing high resolution genetic and physical maps. These entry points will concentrate in regions of the genome that are transcriptionally active. The availability of such STSs can guide the initial genome sequencing efforts towards those genes of which the correlation of their protein products with a defined biological phenomenon is known. The gene sequence will provide the complete primary protein structure, needed as a first step in understanding function; comparison of the upstream sequences of the genes of coregulated proteins may yield identification of specific cis-acting transcriptional regulatory elements. Once a fair part of the genome is sequenced, a shortcut from partial protein sequence to the map position of the gene will exist, immediately correlating an open reading frame with a characterized protein. Eventually, comprehensive databases of proteins and the physical genomic map will completely merge and assist in unraveling the hereditary information stored in the sequence of the complete genome.

We would like to thank Matthias Kretschmer for his collaboration on the 6-phospho-fructo-2-kinase project, Maria Vanderstichelen for helping with the preparation of the manuscript and William Lane for suggesting the use of multiple wave length detection for the identification of tryptophan. A. J. L. was supported by grant DE-FG02-87ER60565 from the Department of Energy to George Church.

Received November 13, 1989

6 References

[1] Hunkapiller, M., Kent, S., Caruthers, M., Dreyer, W., Firca, J., Giffin, C., Horvath, S., Hunkapiller, T., Tempst, P. and Hood, L., *Nature* 1984, *310*, 105–111.

[2] Shively, J. E., Paxton, R. J. and Lee, T. D., *TIBS* 1989, *14*, 246–252.

[3] Tempst, P. and Riviere, L., *Anal. Biochem.* 1989, *183*, 290–300.

[4] Speicher, W. D., in: Hugli, T. E. (Ed.), *Techniques in Protein Chemistry*, Academic Press, San Diego and London, 1989, pp. 24–35.

[5] Aebersold, R. H., Teplow, D. B., Hood, L. E. and Kent, S. B. H., *J. Biol. Chem.* 1986, *261*, 4229–4238.
[6] Vanderkerckhove, J., Bauw, G., Puype, M., Van Damme, J. and Van Montagu, M., *Eur. J. Biochem.* 1985, *152*, 9–19.
[7] Matsudaira, P., *J. Biol. Chem.* 1987, *262*, 10035–10038.
[8] Eckerskorn, C., Mewes, W., Goretzki, H. and Lottspeich, F., *Eur. J. Biochem.* 1988, *176*, 509–519.
[9] Xu, Q-Y. and Shively, J. E., *Anal. Biochem.* 1988, *170*, 19–30.
[10] Walsh, M. J., McDougall, J. and Wittman-Liebold, B., *Biochemistry* 1988, *27*, 6867–6876.
[11] Bauw, G., De Loose, M., Inze, D., Van Montagu, M. and Vandekerckhove, J., *Proc. Natl. Acad. Sci. USA* 1987, *84*, 4806–4810.
[12] Eckerskorn, C., Jungblut, P., Mewes, W., Klose, J. and Lottspeich, F., *Electrophoresis* 1988, *9*, 830–838.
[13] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.
[14] Klose, J., *Humangenetik* 1975, *26*, 211–234.
[15] Celis, J. E. and Bravo, R., *Two-dimensional Gel Electrophoresis*, Academic Press, Orlando, FL 1984.
[16] Dunbar, B. S., *Two-dimensional Electrophoresis and Immunological Techniques*, Plenum Press, New York 1987.
[17] Garrels, J. I. and Franza, B. R., Jr., *J. Biol. Chem.* 1989, *264*, 5299–5312.
[18] Garrels, J. I., *J. Biol. Chem.* 1979, *254*, 7961–7977.
[19] Celis, J. E. (Ed.), *Paper Symposium: Protein Databases in Two-dimensional Electrophoresis*, *Electrophoresis* 1989, *10*, 71–164.
[20] Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. and Kent, S. B. H., *Proc. Nat. Acad. Sci. USA* 1987, *84*, 6970–6974.
[21] Eckerskorn, C. and Lottspeich, F., *Chromatographia* 1990, in press.
[22] Kennedy, T. E., Gawinowicz, M. A., Barzilai, A., Kandel, E. R. and Sweatt, J. D., *Proc. Natl. Acad. Sci. USA* 1988, *85*, 7008–7012.
[23] Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A., *Science* 1988, *239*, 487–491.
[24] Lee, C. C., Wu, X., Gibbs, R. A., Cook, R. G., Muzny, D. M. and Caskey, C. T., *Science* 1988, *239*, 1288–1291.
[25] Ramamgli, L. S. and Rodriguez, L. V., *Electrophoresis* 1985, *6*, 559, 563.
[26] Markwell, M. K., *Anal. Biochem.* 1982, *125*, 427–432.
[27] Laemmli, U. K., *Nature* 1970, *227*, 680–685.
[28] Anderson, L., *Two-Dimensional Electrophoresis: Operation of the ISO-DALT System*, Large Scale Biology Press, Washington, DC 1988.
[29] Anderson, L., *2-D News*, Large Scale Biology Press, Rockville, 1989, No. 3.
[30] Samal, B. B., *Anal. Biochem.* 1987, *163*, 42–44.
[31] Towbin, H., Staehelin, T. and Gordon, J., *Proc. Natl. Acad. Sci. USA* 1979, *76*, 4350–4354.
[32] Szewczyk, B. and Kozloff, L. M., *Anal. Biochem.* 1985, *150*, 403–407.
[33] Hewick, R. M., Hunkapiller, M. W., Hood, L. E. and Dreyer, W. J., *J. Biol. Chem.* 1981, *256*, 7990–7997.
[34] Celis, J. E., Ratz, G. P., Madsen, P., Gesser, B., Lauridsen, J. B., Kwee, S., Rasmussen, H. H., Nielsen, H. V., Crueger, D., Basse, B., Leffers, H., Honore, B., Moller, O., Celis, A., Vandekerchove, J., Bauw, G., Van Damme, J., Puype, M. and Van den Bulcke, M., *FEBS Lett.* 1989, *244*, 247–254.
[35] Bauw, G., Van Damme, J., Puype, M., Vandekerckhove, J., Gesser, B., Ratz, G. P., Lauridsen, J. B. and Celis, J. E., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 7701–7705.
[36] Klose, J., *Electrophoresis* 1989, *10*, 140–152.
[37] Welinder, K. G., *Anal. Biochem.* 1988, *174*, 54–64.
[38] Hermodson, M. A., Ericsson, L. H., Neurath, H. and Walsh, K. A., *Biochemistry* 1973, *12*, 3146–3153.
[39] Brown, J. R. and Hartley, B. S., *Biochemistry* 1966, *101*, 214–228.
[40] Nedkov, P., Oberthur, W. and Braunitzer, G., *Biol. Chem. Hoppe-Seyler* 1985, *366*, 421–430.
[41] Tsunasawa, S., Masaki, T., Hirose, M., Soejima, M. and Sakiyama F., *J. Biol. Chem.* 1989, *264*, 3832–3839.
[42] Friedman, M., Krull, L. H. and Cavins, J. F., *J. Biol. Chem.* 1970, *254*, 3868.
[43] Andrews, P. C. and Dixon, J. E., *Anal. Biochem.* 1987, *161*, 524–528.
[44] Fullmer, C. S., *Anal. Biochem.* 1984, *142*, 336–339.
[45] Amons, R., *FEBS Lett.* 1987, *212*, 68–72.
[46] Brown, J. R., *Fed. Proc.* 1974, *33*, 1389, abstract #941.
[47] Stone, K. L. and Williams, K. R., in: Schlesinger, D. H. (Ed.), *Macromolecular Sequencing and Synthesis*, Alan R. Liss, New York 1988, pp. 7–24.
[48] Stone, K. L., LoPresti, M. B., Williams, N. D., Crawford, J. M. DeAngelis, R. and Williams, K. R. in: Hugli, T. E. (Ed.), *Techniques in Protein Chemistry*, Academic Press, San Diego 1988, pp. 377–391.
[49] Simpson, R. J., Moritz, R. L., Begg, G. S., Rubira, M. R. and Nice, E. C., *Anal. Biochem.* 1989, *177*, 221–236.
[50] Grego, B., Van Driel, I. R., Goding, J. W. Nice, E. C. and Simpson, R. J., *Int. J. Peptide Protein Res.* 1986, *27*, 201–207.
[51] Burgoyne, R., Stacey, C., Young, P., Astephen, N. and Merion, M., in: Hugli, T. E. (Ed.), *Techniques in Protein Chemistry*, Academic Press, San Diego 1988, pp. 399–413.
[52] Laursen, R. A., L'Italien, J. J., Nagarkatti, S. and Miller, D. L., *J. Biol. Chem.* 1981, *256*, 8102–8109.
[53] Maniatis, T., Fritsch, E. F. and Sambrook, J., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory 1982.
[54] Marks, A. R., Tempst, P., Hwang, K. S., Taubman, M. B., Inui, M., Chadwick, C., Fleischer, S. and Nadal-Ginard, B., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 8683–8687.
[55] Olson, M., Hood, L., Cantor, C. and Botstein, D., *Science*, *245*, 1434–1435.
[56] Baringa, M., *Science* 1989, *246*, 32–33.